



Guía docente

Datos Identificativos					2014/15
Asignatura (*)	Lenguajes Naturales		Código	614111625	
Titulación	Enxeñeiro en Informática				
Descritores					
Ciclo	Periodo	Curso	Tipo	Créditos	
1º y 2º Ciclo	1º cuatrimestre	Todos	Optativa	4	
Idioma	CastellanoGallego				
Prerrequisitos					
Departamento	Computación				
Coordinador/a	Vilares Ferro, Jesus	Correo electrónico	jesus.vilares@udc.es		
Profesorado	Vilares Ferro, Jesus	Correo electrónico	jesus.vilares@udc.es		
Web	campusvirtual.udc.es				



Descrición general	<p>DESCRICIÓN:</p> <p>Esta materia aborda o Procesamento da Linguaxe Natural (NLP, Natural Language Processing), a rama das ciencias computacionais encargada do deseño e implementación dos elementos software e hardware necesarios para o tratamento computacional da linguaxe natural, entendendo como tal toda linguaxe humana, en contraposición ás linguaxes formais propias do ámbito lóxico, matemático, ou computacional. O obxectivo último que se persegue, e que aínda está moi lonxe de se conseguir, é o da comprensión da linguaxe humana por parte da computadora.</p> <p>Brevemente, os obxectivos xerais da materia son:</p> <ul style="list-style-type: none">- Diferenciar entre linguaxes naturais e linguaxes artificiais.- Comprender a complexidade da linguaxe humana.- Familiarizarse coa terminoloxía lingüística.- Coñecer os algoritmos, técnicas e métodos máis utilizados actualmente para o tratamento automático dos fenómenos léxicos, sintáticos e semánticos da linguaxe humana.- Asumir que o procesamento da linguaxe natural non se pode automatizar completamente, pero que se poden desenvolver solucións satisfactorias na práctica.- Identificar os aspectos da linguaxe sobre os que debemos traballar máis para obter sistemas de procesamento da linguaxe natural útiles.- Reflexionar sobre os avances realizados no campo e os erros cometidos ao longo das últimas décadas. <p>*CONTEXTUALIZACIÓN:</p> <p>A materia Linguaxes Naturais impártese baixo diversas denominacións como materia optativa de segundo ciclo de EI en diversos plans de estudo, entre eles os das universidades de Alacante (Enxeñaría da Linguaxe Natural, 7.5 créditos LRU), Granada (Interfaces Software en Linguaxe Natural, 6 créditos LRU), País Vasco (Procesamento da Linguaxe Natural, 6 créditos LRU), Politécnica de Madrid (Linguaxe Natural, 6 créditos LRU), Politécnica de Valencia (Linguaxe Natural, 4.5 créditos LRU) e Vigo (Linguaxes Naturais, 6 créditos LRU). Como caso particular, a Universidade Alfonso X O Sabio inclúe dúas materias optativas de primeiro ciclo, denominadas respectivamente Lingüística Computacional I (4,5 créditos LRU) e Lingüística Computacional II (6 créditos LRU) que desenvolven un temario similar.</p> <p>No caso da Facultade de Informática da Universidade da Coruña, optouse por unha materia optativa de segundo ciclo de EI de 4 créditos ECTS (2.5 teóricos e 1.5 prácticos). Esta posición xustifícase pola forte interrelación que mantén con outras materias do plan de estudos, xa que a materia conxuga aspectos da intelixencia artificial, a teoría de autómatas e linguaxes formais, e a construción de compiladores. Este feito permite ao alumno relacionar conceptos, métodos e técnicas que se ensinan en materias dispares. Por outra banda, na materia combínanse os aspectos teóricos, centrados no estudo de certos formalismos, cos aspectos prácticos derivados da realización de algoritmos e programas de computador que transforman eses fundamentos teóricos en aplicacións reais.</p>
---------------------------	---

Competencias de la titulación	
Código	Competencias de la titulación
A1	Aprender de manera autónoma nuevos conocimientos y técnicas avanzadas adecuadas para la investigación, el diseño y el desarrollo de sistemas y servicios informáticos.
A3	Concebir y planificar el desarrollo de aplicaciones informáticas complejas o con requisitos especiales.
A5	Saber especificar, diseñar e implementar sistemas inteligentes cuando las soluciones convencionales no resultan satisfactorias.
A8	Concebir, desplegar, organizar y gestionar un servicio informático complejo.
B1	Aprender a aprender.
B2	Resolver problemas de forma efectiva.
B3	Aplicar un pensamiento crítico, lógico y creativo.
B4	Aprendizaje autónomo.



B5	Trabajar de forma colaborativa.
B6	Comportarse con ética y responsabilidad social como ciudadano y como profesional.
B8	Trabajar en equipos de carácter interdisciplinar.
B9	Capacidad para tomar decisiones.
B10	Capacidad de gestión de la informática (captación y análisis de la información).
B15	Motivación por la calidad.
C3	Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida.
C5	Entender la importancia de la cultura emprendedora y conocer los medios al alcance de las personas emprendedoras.
C6	Valorar críticamente el conocimiento, la tecnología y la información disponible para resolver los problemas con los que deben enfrentarse.
C7	Asumir como profesional y ciudadano la importancia del aprendizaje a lo largo de la vida.
C8	Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad.

Resultados de aprendizaje			
Competencias de materia (Resultados de aprendizaje)	Competencias de la titulación		
Comprender la descripción lingüística de los fenómenos presentes en el lenguaje natural.	A1	B1 B3 B4	
Distinguir entre el nivel léxico, sintáctico y semántico del lenguaje.		B3	
Reconocer los fenómenos del lenguaje que son tratables y aquéllos que no lo son.		B3	
Conocer los algoritmos, técnicas y métodos más empleados actualmente en procesamiento del lenguaje natural.	A1	B1 B4	C3 C6 C8
Comprender la representación formal de diversos fenómenos léxicos, sintácticos y semánticos del lenguaje humano.	A3 A5	B2 B3 B9 B15	C3
Diseñar algoritmos y estructuras de datos para el tratamiento de diversos fenómenos de los lenguajes humanas.	A3 A5	B2 B3 B8 B9 B10 B15	C3 C6
Implementar las soluciones propuestas.	A3 A5	B2 B3 B6 B9 B10 B15	C3 C6
Usar las técnicas y métodos del procesamiento del lenguaje natural para resolver problemas reales de comunicación hombre-máquina.	A1 A3 A5 A8	B2 B3 B4 B9 B10	C3 C6 C8
Asumir la complejidad del lenguaje humano y las limitaciones de los ordenadores para tratarlo.	A5	B3	C6
Aceptar soluciones parciales a un problema ante la imposibilidad de obtener soluciones completas.	A5	B3 B9 B15	C6



Valorar el esfuerzo que requiere realizar avances en un campo complejo.		B6 B15	C6 C7 C8
Rechazar la compartimentación del saber en áreas cerradas (por ejemplo, ciencias en contraposición a humanidades)	A5	B3 B5 B8	C6 C7 C8
Colaborar en el desarrollo de recursos y software libre.			C3 C5 C8

Contenidos	
Tema	Subtema
INTRODUCCIÓN AL PROCESAMIENTO DEL LENGUAJE NATURAL	Niveles de análisis. Ambigüedad.
ANÁLISIS LÉXICO	Segmentación de textos. Morfología flexiva y derivativa. Modelización de grandes diccionarios. Autómatas finitos acíclicos deterministas numerados. Traductores de estado finito y morfología de dos niveles.
ETIQUETACIÓN	Modelos de Markov ocultos. Ejecución eficiente de los modelos de Markov ocultos. Técnicas de suavizado. Tratamiento de palabras desconocidas. Aprendizaje de etiquetas basado en transformaciones y dirigido por el error.
ANÁLISIS SINTÁCTICO: GRAMÁTICAS INDEPENDIENTES DEL CONTEXTO	Esquemas de análisis sintáctico. Análisis ascendente. El algoritmo de Earley. Autómatas a pila y programación dinámica. Análisis sintáctico LR generalizado. Representación compartida de los árboles de análisis sintáctico. Análisis sintáctico probabilístico.
ANÁLISIS SINTÁCTICO: GRAMÁTICAS SUAVEMENTE DEPENDIENTES DEL CONTEXTO	Gramáticas de adjunción de árboles. Análisis sintáctico de gramáticas de adjunción de árboles. Autómatas para las gramáticas de adjunción de árboles. Representación compartida de los árboles de derivación. Gramáticas de adjunción de árboles probabilísticas.
ANÁLISIS SEMÁNTICO	Estructuras de rasgos y formalismos basados en unificación. Relaciones léxicas: WordNet y EuroWordNet.
RECUPERACIÓN DE INFORMACIÓN (RI)	Conceptos básicos Modelos de recuperación: booleano, vectorial y probabilístico Los procesos de indexación y recuperación Evaluación en RI RI sobre web. Un caso práctico: Google Aplicaciones del procesamiento del lenguaje natural a RI: la variación lingüística
EXTRACCIÓN DE INFORMACIÓN (EI)	Conceptos básicos Arquitectura de un sistema de EI Tareas de EI Evaluación en EI Ejemplos de sistemas de EI: FASTUS y otros



BÚSQUEDA DE RESPUESTAS (BR)	<p>Conceptos básicos</p> <p>BR vs. RI/EI</p> <p>Arquitectura de un sistema de BR</p> <p>Procesamiento de la pregunta</p> <p>Recuperación y selección de documentos/pasajes</p> <p>Extracción de la respuesta</p> <p>Evaluación en PR</p>
TRADUCCIÓN AUTOMÁTICA (TA)	<p>Conceptos básicos y problemática</p> <p>Técnicas "clásicas"</p> <p>Técnicas estadísticas</p> <p>Aplicaciones en RI interlingüe</p>

Planificación			
Metodologías / pruebas	Horas presenciales	Horas no presenciales / trabajo autónomo	Horas totales
Prueba de respuesta breve	3	4	7
Sesión magistral	17	17	34
Taller	8	8	16
Prácticas de laboratorio	10	20	30
Mesa redonda	4	4	8
Atención personalizada	5	0	5

(*Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos

Metodologías	
Metodologías	Descripción
Prueba de respuesta breve	A final de curso se realizará una prueba escrita con problemas de similar dificultad a los planteados durante el curso.
Sesión magistral	En las clases presenciales de teoría, el profesor realizará una breve descripción de los contenidos temáticos y de los objetivos básicos perseguidos, con el fin de dotar al alumno de una visión global de la materia. Además tratará de establecer interrelaciones con otros conceptos previamente adquiridos, de forma que se pueda establecer una línea temporal, y expondrá la bibliografía recomendada. Seguidamente pasará a desarrollar los contenidos teóricos, utilizando como método la clase magistral.
Taller	En las clases presenciales de problemas, con el fin de afianzar los conceptos teóricos, se presentarán supuestos prácticos, que en un principio serán resueltos por el profesor para orientar a los alumnos. Como actividades no presenciales, se expondrán ejercicios adicionales que el alumno deberá resolver y comentar/corregir con el profesor durante las horas de clases prácticas. Se trata de fomentar la participación de los alumnos y promover, en la medida del posible, el diálogo abierto y la valoración de soluciones.
Prácticas de laboratorio	Las clases de prácticas de laboratorio obligan a la implementación de soluciones para un problema dado. Se impondrá una periodicidad en su entrega para fomentar el estudio continuo. El enunciado de las prácticas, que se proporcionará con antelación suficiente, detallará el problema y las especificaciones, que el alumno deberá respetar estrictamente. Este enunciado se analizará con los alumnos. Posteriormente, la labor del profesor será la de supervisar las sesiones de prácticas, solucionando dudas, corrigiendo errores de interpretación, sugiriendo lecturas, etc.
Mesa redonda	Se fomentará el espíritu crítico de los alumnos y el desarrollo de su capacidad para argumentar mediante la realización de debates y discusiones en forma de tutorías colectivas.

Atención personalizada	
Metodologías	Descripción



Prácticas de laboratorio Taller Sesión magistral	La labor del profesor será, tanto en las clases teóricas como prácticas, la de supervisar el trabajo del alumno, solucionando dudas, corrigiendo errores de interpretación, sugiriendo lecturas, etc., no sólo como grupo, sino también como individuo.
--	---

Evaluación		
Metodologías	Descripción	Calificación
Prueba de respuesta breve	Prueba escrita con problemas de dificultad semejante a los expuestos durante lo curso. Se evaluarán el dominio de conocimientos teóricos y su aplicación en resolución de problemas. El examen final es optativo, por lo que el alumno puede optar por incrementar el peso de la parte práctica hasta el 80% de la nota de la materia.	40
Mesa redonda	Se valorará la participación en las clases y tutorías colectivas.	3
Prácticas de laboratorio	Se presentarán una o varias prácticas a realizar individualmente o en grupo, según el alcance de las mismas. Se evaluarán la puesta en práctica de los conocimientos de la materia, la metodología de diseño, el funcionamiento y la innovación. Independientemente de la nota obtenida en el resto de los apartados de evaluación, el alumno deberá superar esta parte práctica para superar la materia. El examen final es optativo, por lo que el alumno puede optar por incrementar el peso de la parte práctica hasta el 80% de la nota de la materia.	40
Taller	Se valorará la participación en las clases y tutorías colectivas.	7
Sesión magistral	Se valorará la participación en las clases y tutorías colectivas.	10
Otros		

Observaciones evaluación
En el caso de nuevos alumnos, al no haber ya ni clases teóricas ni horario de laboratorios asignados a la materia, la calificación de la materia se basará únicamente en la nota obtenida en el examen, del tipo de respuesta breve.

Fuentes de información



<p>Básica</p>	<ul style="list-style-type: none"> - Christopher D. Manning y Hinrich Schütze (1999). Foundations of Statistical Natural Language Processing. The MIT Press, Cambridge (Massachusetts, EE.UU.)/Londres (Reino Unido) - Robert Dale, Hermann Moisi y Harold Somers (eds.) (2000). Handbook of Natural Language Processing. Marcel Dekker, Inc., Nueva York/Basilea - Christopher D. Manning, Prabhakar Raghavan, y Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press, Cambridge - A. Arampatzis, Th. P. van der Weide, P. van Bommel y C.H.A. Koster (2000). Linguistically-motivated Information Retrieval. En Vol. 69 de Encyclopedia of Library and Information Science, pág. 201-222. Marcel Dekker - Peter Jackson y Isabelle Moulinier (2007). Natural language processing for online applications : text retrieval, extraction and categorization (2nd ed.). John Benjamins, Amsterdam/Philadelphia - James Allen (1995). Natural Language Understanding (2a ed.). The Benjamin/Cummings Publishing Company, Inc., Redwood City, CA, EE.UU - Marius Pasca (2003). Open-domain question answering from large text collections . CSLI Publications, Standford - W. Bruce Croft, Donald Metzler y Trevor Strohman (2009). Search Engines: Information Retrieval in Practice. Pearson Education, Upper Saddle River, NJ, USA - Daniel Jurafsky y James H. Martin (2009). Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (2nd ed.). Pearson Prentice Hall, Upper Saddle River, New Jersey, EE.UU
<p>Complementaría</p>	<ul style="list-style-type: none"> - Pierre M. Nugues (2006). An introduction to Language Processing with Perl and Prolog. Springer , Berlin - Heidelberg- New York - W. John Hutchings y Harold L. Somers (1992). An Introduction to Machine Translation. Academic Press, Londres/San Diego - Gregory Grefenstette (ed.) (1998). Cross-language information retrieval. Kluwer Academic Publishers, Boston - José Francisco Quesada Moreno y José Gabriel De Amores Carredano (2000). Diseño e implementación de sistemas de traducción automática. Secretariado de Publicaciones de la Universidad de Sevilla - Piek Vossen (ed.) (1998). EuroWordNet. A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht, Holanda - Eric Wehrli (1997). L?analyse syntaxique des langues naturelles. Masson, París - Ricardo Baeza-Yates y Berthier Ribeiro-Neto (1999). Modern Information Retrieval. Addison Wesley y ACM Press, Harlow, Inglaterra - Steven Bird, Ewan Klein y Edward Loper (2009). Natural Language Processing with Python. O'Reilly Media, Sebastopol, USA - Klaas Sikkel (1997). Parsing Schemata ? A Framework for Specification and Analysis of Parsing Algorithms, Texts in Theoretical Computer Science ? An EATCS Series. Springer-Verlag, Berlín/Heidelberg/Nueva York - Fernando C. N. Pereira y Stuart M. Shieber (1987). Prolog and Natural Language Analysis, CSLI Lecture Notes 10. Center for the Study of Language and Information, Stanford, CA, EE.UU. - José Luis Vicedo González (2003). Recuperación de Información de alta precisión: Los sistemas de búsqueda de respuestas. Sociedad Española para el Procesamiento del Lenguaje Natural - Jerry R. Hobbs (1993). The generic information extraction system. En Proceedings of the 5th Conference on Message understanding (MUC-5), pág. 87-91. Morgan Kauffman Publishers, San Francisco, USA

Recomendaciones

Asignaturas que se recomienda haber cursado previamente

Asignaturas que se recomienda cursar simultáneamente

Asignaturas que continúan el temario

Teoría de Autómatas y Lenguajes Formal/614111301



Otros comentarios

(*) La Guía Docente es el documento donde se visualiza la propuesta académica de la UDC. Este documento es público y no se puede modificar, salvo cosas excepcionales bajo la revisión del órgano competente de acuerdo a la normativa vigente que establece el proceso de elaboración de guías